Solvency Games
How to gamble forever

Nevin Kapur
with Noam Berger, Leonard Schulman, and Vijay Vazirani
Main theme

▶ Many investors want to afford a comfortable existence and minimize the risk of catastrophic disruption to this way of life.
▶ Two investment goals
  ▶ maximizing profit
  ▶ minimizing risk
are related but distinct.
Example

You have $50 and a choice of two investments:

A  lose $3 w.p. 0.5; gain $4 w.p. 0.5
B  lose $30 w.p. 0.5; gain $40 w.p. 0.5

Investment B has higher expected profit, but ....
Example

You have $50 and a choice of two investments:

- A lose $3 w.p. 0.5; gain $4 w.p. 0.5
- B lose $30 w.p. 0.5; gain $40 w.p. 0.5

Investment B has higher expected profit, but . . . .

What about more subtle choices?

- A lose $3 w.p. 0.5; gain $4 w.p. 0.5
- C lose $4 w.p. 0.4; gain $3 w.p. 0.6
Your wealth is an integer \( w > 0 \).
Framework

- Your wealth is an integer \( w > 0 \).
- You have a choice of several (finite) “investments” A, B, C, ....
- Each investment is a probability distribution on integers with finite support. E.g.,
  \[
  A \begin{cases}
    q^A_{-3} = 0.5, & q^A_{+4} = 0.5 \\
    \end{cases}
  \]
  \[
  C \begin{cases}
    q^C_{-4} = 0.4, & q^C_{+3} = 0.6 \\
    \end{cases}
  \]
Framework

- Your wealth is an integer $w > 0$.
- You have a choice of several (finite) “investments” A, B, C, . . . .
- Each investment is a probability distribution on integers with finite support. E.g.,
  \[ A \{ q^A_{-3} = 0.5, q^A_{+4} = 0.5 \} \]
  \[ C \{ q^C_{-4} = 0.4, q^C_{+3} = 0.6 \} \]
- At each discrete time unit, you pick an investment and add a sample $\delta$ from the its distribution to your wealth: $w \leftarrow w + \delta$
- You keep playing until $w \leq 0$. 
Framework

- Your wealth is an integer $w > 0$.
- You have a choice of several (finite) "investments" A, B, C, . . . .
- Each investment is a probability distribution on integers with finite support. E.g.,
  \[
  A \{ q_{-3}^A = 0.5, q_{+4}^A = 0.5 \} \\
  C \{ q_{-4}^C = 0.4, q_{+3}^C = 0.6 \}
  \]
- At each discrete time unit, you pick an investment and add a sample $\delta$ from its distribution to your wealth: $w \leftarrow w + \delta$
- You keep playing until $w \leq 0$.

How should you pick investments so as to minimize the probability of every going broke?
Strategies

A strategy or policy or decision rule is map from wealth to investments $\pi : \mathbb{Z}_+ \rightarrow \{A, B, C, \ldots\}$. 
Strategies

A *strategy* or *policy* or *decision rule* is map from wealth to investments $\pi : \mathbb{Z}_+ \to \{A, B, C, \ldots \}$.

$v^\pi(w)$: probability of ever going broke starting from wealth $w$ under policy $\pi$

**Goal:** find a strategy $\pi^*$ that is *optimal*, i.e., $v^{\pi^*}(w) \leq v^{\pi}(w)$ for all $w$ and $\pi$.
A more general framework

- \( S \): set of states
- \( A_s \): set of actions for state \( s \in S \)
- \( p(\cdot \mid s, a) \): transition probability function on the set of states
- \( r(s, a) \): (expected) reward function
- \( d : S \rightarrow A_s \): decision rule
A more general framework

- $S$: set of states
- $A_s$: set of actions for state $s \in S$
- $p(\cdot \mid s, a)$: transition probability function on the set of states
- $r(s, a)$: (expected) reward function
- $d : S \to A_s$: decision rule

At each discrete time the system occupying a state $s \in S$ chooses an action $d(s) \in A_s$ and makes a transition according to $p(\cdot \mid s, a)$ collecting a reward $r(s, a)$.

This is a Markov Decision Process (MDP).
Markov Decision Problem

- $X_t$: state of the system at time $t$
- $Y_t := d(X_t)$: action taken

The *expected total reward* is defined as

$$v_d(s) := \lim_{N \to \infty} E_s^d \left\{ \sum_{t=1}^{N} r(X_t, Y_t) \right\},$$

where $E_s^d$ represents taking expectation with $X_1 = s$ using decision rule $d$.

**Goal:** find a decision rule $d^*$ that is optimal, i.e., $v_{d^*}(s) \geq v_d(s)$ for all $s \in S$ and decision rules.
Our MDP

\[ S = \{1, 2, \ldots\} \cup \{0\} \]
\[ A_s = \{A, B, \ldots\} \text{ for } s \in \{1, 2, \ldots\} \text{ and } A_0 \text{ is the singleton action that stays in the current state w.p. 1.} \]
\[ p(j | s, A) = q^A_{j-s} \]
\[ r(s, a) = - \sum_{j \in S} 1\{s \in \mathbb{Z}_+ \text{ and } j = 0\} p(j | s, a) \]
Our MDP

\[ S = \{1, 2, \ldots \} \cup \{0\} \]

\[ A_s = \{A, B, \ldots \} \text{ for } s \in \{1, 2, \ldots \} \text{ and } A_0 \text{ is the singleton action that stays in the current state w.p. 1.} \]

\[ p(j \mid s, A) = q_{j-s}^A \]

\[ r(s, a) = - \sum_{j \in S} 1\{s \in \mathbb{Z}_+ \text{ and } j = 0\} p(j \mid s, a) \]

Given a decision rule, the expected total reward is just the negative of the probability of every going broke.
Existence of optimal strategy

Fact
\[ r(s, a) \leq 0 \quad + \quad \text{discrete state space} \quad + \quad \text{finite action set} \quad \implies \quad \text{there exists an optimal decision rule} \]
Existence of optimal strategy

Fact
\[ r(s,a) \leq 0 \quad + \quad \text{discrete state space} \quad + \quad \text{finite action set} \quad \implies \quad \text{there exists an optimal decision rule} \]

- Result holds even if one is allowed to use randomized strategies and change strategies with time units and use history.
Existence of optimal strategy

Fact
\[ r(s, a) \leq 0 + \text{discrete state space} + \text{finite action set} \implies \text{there exists an optimal decision rule} \]

- Result holds even if one is allowed to use randomized strategies and change strategies with time units and use history.

We’ll focus on the structure of optimal strategies and algorithms for determining them.
Pure strategies

What’s the probability of going broke when one employs a *pure strategy* (one that only uses the same investment at every wealth)? [Gambler’s Ruin]
Pure strategies

What’s the probability of going broke when one employs a pure strategy (one that only uses the same investment at every wealth)?

**[Gambler’s Ruin]**

$v(w)$: probability of ever going broke starting at $w$ using only investment ($q_{-l}, \ldots, q_{+r}$)
Pure strategies

What’s the probability of going broke when one employs a pure strategy (one that only uses the same investment at every wealth)?

[Gambler’s Ruin]

$v(w)$: probability of ever going broke starting at $w$ using only investment $(q_{-l}, \ldots, q_{+r})$

Then, for $w \geq 1$,

$$v(w) = \sum_{j=-l}^{r} q_{j} v(w + j)$$

and $v(0) = \cdots = v(-l + 1) = 1$. 

Pure strategies

What’s the probability of going broke when one employs a pure strategy (one that only uses the same investment at every wealth)?

[Gambler’s Ruin]

$v(w)$: probability of ever going broke starting at $w$ using only investment $(q_{-l}, \ldots, q_{+r})$

Then, for $w \geq 1$,

$$v(w) = \sum_{j=-l}^{r} q_j v(w + j)$$

and $v(0) = \cdots = v(-l + 1) = 1$.

The characteristic polynomial of this recurrence is

$$q(z) := z^l - \sum_{j=-l}^{r} q_j z^{j+l}.$$
Positive drift: $\sum_{j=-l}^{l} jq_j > 0 \iff v(w) < 1$ for $w \geq 1$
Positive drift: $\sum_{j=-l}^{r} jq_j > 0 \iff v(w) < 1$ for $w \geq 1$

- A simple root $\lambda$ of the characteristic polynomial $q(z)$ contributes $c\lambda^w$ to $v(w)$. 

Theorem: positive drift $\Rightarrow 1$.

Proof. Rouché's theorem + continuity of zeros.
Positive drift: $\sum_{j=-l}^{r} j q_j > 0 \iff v(w) < 1$ for $w \geq 1$

- A simple root $\lambda$ of the characteristic polynomial $q(z)$ contributes $c\lambda^w$ to $v(w)$.
- The characteristic polynomial $q(z)$ has degree $l + r$.
- We only have $l$ initial conditions.
Positive drift: $\sum_{j=-l}^{r} j q_j > 0 \iff v(w) < 1$ for $w \geq 1$

- A simple root $\lambda$ of the characteristic polynomial $q(z)$ contributes $c\lambda^w$ to $v(w)$.
- The characteristic polynomial $q(z)$ has degree $l + r$.
- We only have $l$ initial conditions.
- Since $v(w) \to 0$ as $w \to \infty$, it must be that if $|\lambda| \geq 1$, it cannot have a nonzero contribution to $v(w)$. 
Positive drift: $\sum_{j=-l}^{r} jq_j > 0 \iff v(w) < 1$ for $w \geq 1$

- A simple root $\lambda$ of the characteristic polynomial $q(z)$ contributes $c\lambda^w$ to $v(w)$.
- The characteristic polynomial $q(z)$ has degree $l + r$.
- We only have $l$ initial conditions.
- Since $v(w) \to 0$ as $w \to \infty$, it must be that if $|\lambda| \geq 1$, it cannot have a nonzero contribution to $v(w)$.

**Theorem**

**positive drift** $\implies$

1. $q(z)$ has precisely $l$ roots in the interior of the unit disk
2. $q(z)$ has a unique positive root $\lambda_+ < 1$

**Proof.**

Rouché’s theorem + continuity of zeros.
Another point of view

\( \alpha_j \): probability that starting from wealth \( w \), the \textit{first crossing} to the left of \( w \) is to \( w - j \), \( (j = 1, \ldots, l) \)
Another point of view

\( \alpha_j \): probability that starting from wealth \( w \), the *first crossing* to the left of \( w \) is to \( w - j \), \( (j = 1, \ldots, l) \)

Then, for \( w \geq 1 \),

\[
\nu(w) = \sum_{j=1}^{l} \alpha_j \nu(w - j)
\]

and \( \nu(0) = \cdots = \nu(-l + 1) = 0 \).
Another point of view

\( \alpha_j \): probability that starting from wealth \( w \), the \textit{first crossing} to the left of \( w \) is to \( w - j \), \( (j = 1, \ldots, l) \)

Then, for \( w \geq 1 \),

\[
\nu(w) = \sum_{j=1}^{l} \alpha_j \nu(w - j)
\]

and \( \nu(0) = \cdots = \nu(-l + 1) = 0 \).

Its characteristic polynomial is

\[
c(z) = z^l - \sum_{j=1}^{l} \alpha_j z^{l-j},
\]

with \textit{companion matrix} \ldots
C := \[
\begin{pmatrix}
1 & 1 & \cdots & 1 \\
\alpha_l & \alpha_l - 1 & \cdots & \alpha_1
\end{pmatrix},
\]

Further assumption: gcd of support of the investment is 1 (investment is irreducible) ⇒
1. λ+ is algebraically simple
2. λ+ is the unique eigenvalue of maximum modulus irreducible investment with positive drift =
\[v(w) \approx \lambda w + \cdots\]
The problem

Pure strategies

Some details

\[ \mathbf{C} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_l & \alpha_{l-1} & \cdots & \alpha_1 \end{bmatrix}, \]

- \( l \) roots of \( q(z) \) inside the unit disk \( \equiv l \) roots of \( c(z) \) \( \equiv l \) eigenvalues of \( \mathbf{C} \)
- (allows computations of the \( \alpha_j \)'s)
The problem

Pure strategies

Pure strategies

Algorithms

Some details

\[
C := \begin{bmatrix}
1 & 1 & & \\
& 1 & \ddots & \\
\alpha_l & \alpha_{l-1} & \cdots & 1
\end{bmatrix},
\]

- \( l \) roots of \( q(z) \) inside the unit disk \( \equiv l \) roots of \( c(z) \) \( \equiv l \) eigenvalues of \( C \)

- (allows computations of the \( \alpha_j \)'s)

- **Further assumption**: gcd of support of the investment is 1 (investment is *irreducible*)

  1. \( \lambda_+ \) is algebraically simple
  2. \( \lambda_+ \) is the unique eigenvalue of maximum modulus
The problem

Pure strategies

Pure strategies

Algorithms

Some details

\[ C := \begin{bmatrix} 1 & 1 \\ \alpha_l & \alpha_{l-1} & \cdots & \alpha_1 \end{bmatrix}, \]

- \( l \) roots of \( q(z) \) inside the unit disk \( \equiv \) \( l \) roots of \( c(z) \) \( \equiv \)
- \( l \) eigenvalues of \( C \)
- (allows computations of the \( \alpha_j \)'s)
- Further assumption: gcd of support of the investment is 1 (investment is irreducible) \( \implies \)
  1. \( \lambda_+ \) is algebraically simple
  2. \( \lambda_+ \) is the unique eigenvalue of maximum modulus

irreducible investment with positive drift \( \implies \) \( v(w) \simeq \lambda_+^w \)
Optimal strategies

Theorem

finite set of investments + irreducible investment $A$ such that $\lambda_A^+ < \lambda_B^+$ for all positive drift $B$’s $\implies$ any optimal strategy has a pure-$A$ tail
Optimal strategies

Theorem

finite set of investments $\rightarrow$ irreducible investment $A$ such that $\lambda^A_+ < \lambda^B_+$ for all positive drift $B$’s $\implies$ any optimal strategy has a pure-$A$ tail

A lose $3$ w.p. $0.5$; gain $4$ w.p. $0.5$ $\lambda^A_+ \approx .921$
B lose $30$ w.p. $0.5$; gain $40$ w.p. $0.5$ $\lambda^B_+ \approx .992$
C lose $4$ w.p. $0.4$; gain $3$ w.p. $0.6$ $\lambda^C_+ \approx .967
Theorem
\[ \lambda^A_+ = \lambda^B_+ < 1 \text{ and } A \text{ and } B \text{ have eigenvalues (in the unit disk) that are not shared by the other} \implies \text{any strategy with a pure-A (or pure-B) tail is not optimal} \]
Converse

Theorem
\( \lambda^A_+ = \lambda^B_+ < 1 \) and A and B have eigenvalues (in the unit disk) that are not shared by the other \( \implies \) any strategy with a pure-A (or pure-B) tail is not optimal

\( l_A = l_B = 2 \)

Any optimal strategy has tail ABAB\ldots
Converse

Theorem
\[ \lambda^A_+ = \lambda^B_+ < 1 \text{ and } A \text{ and } B \text{ have eigenvalues (in the unit disk) that are not shared by the other} \implies \text{any strategy with a pure-A (or pure-B) tail is not optimal} \]

\[ l_A = l_B = 2 \]
Any optimal strategy has tail \( ABAB \ldots \).

Open problem
Is the optimal tail always periodic?
Strategies with pure tails

If $\pi$ has a pure-A tail beyond $M$ then for $w > M$

$$v(w) = \sum_{j=1}^{I_A} \alpha_j v(w - j).$$

(Recall that $\alpha_j$ is the probability of first crossing to the left of $w$ being at $w - j$.)
Strategies with pure tails

If $\pi$ has a pure-A tail beyond $M$ then for $w > M$

$$v(w) = \sum_{j=1}^{l_A} \alpha_j v(w - j).$$

(Recall that $\alpha_j$ is the probability of first crossing to the left of $w$ being at $w - j$.)

► We can convert an infinite problem into a finite problem.

Here $r := \max_B \{r_B\}$ and $\alpha_+ := \sum_{j=1}^{l_A} \alpha_j$. 
A finite MDP

- $S = \{1, 2, \ldots, M + r\} \cup \{0, \infty\}$
A finite MDP

- $S = \{1, 2, \ldots, M + r\} \cup \{0, \infty\}$
- $A_s = \{A, B, \ldots\}$ for $s \in \{1, 2, \ldots, M\}$
- $A_s = \{A\}$ for $s \in \{M + 1, \ldots, M + r\}$
- $A_s$ is the singleton action that stays in the current state w.p. 1 for $s \in 0, \infty$
A finite MDP

- \( S = \{1, 2, \ldots, M + r\} \cup \{0, \infty\} \)
- \( A_s = \{A, B, \ldots\} \) for \( s \in \{1, 2, \ldots, M\} \)
- \( A_s = \{A\} \) for \( s \in \{M + 1, \ldots, M + r\} \)
- \( A_s \) is the singleton action that stays in the current state w.p. 1 for \( s \in 0, \infty \)
- \( p(j \mid s, B) = q^B_{j-s} \) for \( s \in \{1, \ldots, M\} \)
- \( p(j \mid s, A) = \alpha_{s-j} \) and \( p(\infty \mid s, A) = 1 - \alpha_+ \) for \( s \in \{M + 1, \ldots, M + r\} \)
A finite MDP

- \( S = \{1, 2, \ldots, M + r\} \cup \{0, \infty\} \)
- \( A_s = \{A, B, \ldots\} \) for \( s \in \{1, 2, \ldots, M\} \)
- \( A_s = \{A\} \) for \( s \in \{M + 1, \ldots, M + r\} \)
- \( A_s \) is the singleton action that stays in the current state w.p. 1 for \( s \in 0, \infty \)
- \( p(j \mid s, B) = q^B_{j-s} \) for \( s \in \{1, \ldots, M\} \)
- \( p(j \mid s, A) = \alpha_{s-j} \) and \( p(\infty \mid s, A) = 1 - \alpha_+ \) for \( s \in \{M + 1, \ldots, M + r\} \)
- \( r(s, a, j) = 1 \) if \( j = \infty \) and \( r(s, a) = \alpha_{\infty} \mathbf{1}[s \in \{M + 1, \ldots, M + r\}] \)
A finite MDP

- \( S = \{1, 2, \ldots, M + r\} \cup \{0, \infty\} \)
- \( A_s = \{A, B, \ldots\} \) for \( s \in \{1, 2, \ldots, M\} \)
- \( A_s = \{A\} \) for \( s \in \{M + 1, \ldots, M + r\} \)
- \( A_s \) is the singleton action that stays in the current state w.p. 1 for \( s \in 0, \infty \)
- \( p(j \mid s, B) = q_{j-s}^B \) for \( s \in \{1, \ldots, M\} \)
- \( p(j \mid s, A) = \alpha_{s-j} \) and \( p(\infty \mid s, A) = 1 - \alpha_+ \) for \( s \in \{M + 1, \ldots, M + r\} \)
- \( r(s, a, j) = 1 \) if \( j = \infty \) and
  \[ r(s, a) = \alpha_\infty 1[s \in \{M + 1, \ldots, M + r\}] \]

Given a decision rule \( d \) (a choice of actions at states \( 1, \ldots, M \)), the expected total reward \( v_d(s) \) is the complement of ever going broke starting from \( s \).
Standard algorithms for solving (finite) MDPs:

1. Value iteration: an iterative scheme that converges to the optimal reward
2. Policy iteration: an iterative scheme that converges in finite number of steps to the optimal policy (decision rule)
3. Linear programming
Value iteration

1. Set $v^0(s) = 0$ for each $s \in S$.
2. For each $s \in S$, compute $v^{n+1}(s)$ using

$$v^{n+1}(s) = \max_{a \in A_s} \left\{ r(s, a) + \sum_{j \in S} p(j \mid s, a)v^n(j) \right\}$$

Fact

$v_k(s) \to v^*(s)$, the optimal probability of continuing forever.

Coming up . . . rates of convergence
Value iteration

1. Set $v^0(s) = 0$ for each $s \in S$.
2. For each $s \in S$, compute $v^{n+1}(s)$ using

$$v^{n+1}(s) = \max_{a \in A_s} \left\{ r(s, a) + \sum_{j \in S} p(j \mid s, a) v^n(j) \right\}$$

Fact

$v^k(s) \to v^*(s)$, the optimal probability of continuing forever

Coming up . . .

rates of convergence
Policy iteration

1. Set $n = 0$ and select an arbitrary decision rule $d_0$.

2. [Policy evaluation] Compute the expected total reward $\{v^n(s)\}_{s \in S}$ for the rule $d_n$ by solving the linear system of equations:

$$v(s) = r(s, d_n(s)) + \sum_{j \in S} p(j \mid s, d_n(s)) v(j), \quad s \in S.$$

3. [Policy improvement] For each $s \in S$, choose $d_{n+1}(s)$ such that

$$d_{n+1}(s) \in \arg \max_{a \in A_s} \left\{ r(s, a) + \sum_{j \in S} p(j \mid s, a) v^n(j) \right\},$$

choosing $d_{n+1}(s) = d_n(s)$ whenever possible.

4. If $d_{n+1}(s) = d_n(s)$ for all $s \in S$, then stop, setting $d^*$ to $d_n$. 
Theorem

\[ \text{local optimality} \implies \text{global optimality}, \text{ so policy iteration terminates in finite number of steps with the optimal policy} \]
Linear programming

Choose $\beta_1, \ldots, \beta_{M+r} > 0$. The optimal value is the solution to the following linear programming:

$$\text{minimize} \sum_{j \in S} \beta_j v(j)$$

subject to

$$v(s) - \sum_{j \in S} p(j \mid s, a)v(j) \geq r(s, a); \quad a \in A_s, \ s \in S,$$

$$v(s) \geq 0; \quad s \in S.$$
Choose $\beta_1, \ldots, \beta_{M+r} > 0$. The optimal value is the solution to the following linear programming:

$$\text{minimize } \sum_{j \in S} \beta_j v(j)$$

subject to

$$v(s) - \sum_{j \in S} p(j \mid s, a) v(j) \geq r(s, a); \quad a \in A_s, s \in S,$$

$$v(s) \geq 0; \quad s \in S.$$  

To determine the optimal policy, one considers the dual.
The dual

maximize \sum_{s \in S} \sum_{a \in A_s} r(s, a)x(s, a)

subject to

\sum_{a \in A_j} x(j, a) - \sum_{s \in S} \sum_{a \in A_s} p(j \mid s, a)x(s, a) \leq \beta_j; \quad j \in S,

x(s, a) \geq 0; \quad a \in A_s, s \in S.
The dual

\[
\text{maximize } \sum_{s \in S} \sum_{a \in A_s} r(s, a)x(s, a)
\]

subject to

\[
\sum_{a \in A_j} x(j, a) - \sum_{s \in S} \sum_{a \in A_s} p(j \mid s, a)x(s, a) \leq \beta_j; \quad j \in S,
\]

\[
x(s, a) \geq 0; \quad a \in A_s, s \in S.
\]

Given an optimal basic feasible solution \(x^*\) to the dual, an optimal decision rule can be determined as

\[
d^*(s) = \begin{cases} 
  a & \text{if } x^*(s, a) > 0 \text{ and } s \in S^* \\
  \text{arbitrary} & \text{otherwise.}
\end{cases}
\]

Here \(S^* := \{s \in S^* : \sum_{a \in A_s} x^*(s, a) > 0\}\).
Convergence rates of value iteration

Theorem

Let $v^*$ denote the optimal probability of continuing forever and $v^n$ the $n$th iterate of value iteration.

Then $v^n \geq u^n$, where for some vector norm $\| \cdot \|$ and $n \geq 1$,

$$\| v^* - u^n \| \leq \delta \| v^* - u^{n-1} \|$$

and $\delta < 1$. 


Convergence rates of value iteration

Theorem

Let $v^*$ denote the optimal probability of continuing forever and $v^n$ the $n$th iterate of value iteration.

- Then $v^n \geq u^n$, where for some vector norm $\| \cdot \|$ and $n \geq 1$,

$$
\| v^* - u^n \| \leq \delta \| v^* - u^{n-1} \|
$$

and $\delta < 1$.

- If in addition all the investments have positive drift then

$$
\delta \leq \max \left\{ 1 - \min_{\text{investment } B} \left\{ \frac{q^B (\lambda + \epsilon)}{(\lambda + \epsilon) l_B} \right\}, \sum_{j=1}^{l_A} \alpha_j (\lambda + \epsilon)^{-j} \right\}
$$

where $\lambda := \max_{\text{investment } B} \lambda_B^+ \text{ and } \epsilon > 0 \text{ is arbitrarily small.}$
Rest of this talk

- Existence of pure tail
- Convergence rate of value iteration
Structure of optimal strategies

Main technical tool: notion of (sub-/super-) harmonicity

Definition
A sequence \((a_n)_{n \geq 1}\) is said to be

harmonic w.r.t. policy \(\pi\) if for all \(n \geq 1\),

\[
a_n = \sum_{j=-l_{\pi(n)}}^{r_{\pi(n)}} q_j^{\pi(n)} a_{n-j} := E_n^{\pi(n)}(a)
\]
Structure of optimal strategies

Main technical tool: notion of (sub-/super-) harmonicity

Definition
A sequence \((a_n)_{n \geq 1}\) is said to be

harmonic w.r.t. policy \(\pi\) if for all \(n \geq 1\),

\[
a_n = \sum_{j=-l_\pi(n)}^{r_\pi(n)} q_j^{\pi(n)} a_{n-j} := E^{\pi(n)}_n(a)
\]

subharmonic w.r.t. policy \(\pi\) if for all \(n \geq 1\),

\[
a_n \leq E^{\pi(n)}_n(a)
\]

superharmonic w.r.t. policy \(\pi\) if for all \(n \geq 1\),

\[
a_n \geq E^{\pi(n)}_n(a)
\]
Key lemma

Lemma
If $a_n, v_n \to 0$ and

1. $(v_n)_{n \geq 1}$ is harmonic w.r.t. policy $\pi$
2. $(a_n)_{n \geq 1}$ is subharmonic (superharmonic) w.r.t. $\pi$
3. $a_n = v_n$ for $n \leq 0$

Then $a_n \leq v_n$ ($a_n \geq v_n$) for all $n$. 

Proof. Look at $b_n := a_n - v_n$.

$b_n \to 0$, $b_n = 0$ for $n \leq 0$, and $b_n$ is subharmonic w.r.t $\pi$.

If $b_n > 0$ for some $n$ then it achieves a maximum somewhere.

The last maximum cannot be subharmonic.
Key lemma

Lemma
If $a_n, v_n \to 0$ and

1. $(v_n)_{n \geq 1}$ is harmonic w.r.t. policy $\pi$
2. $(a_n)_{n \geq 1}$ is subharmonic (superharmonic) w.r.t. $\pi$
3. $a_n = v_n$ for $n \leq 0$

Then $a_n \leq v_n$ ($a_n \geq v_n$) for all $n$.

Proof.

- Look at $b_n := a_n - v_n$
- $b_n \to 0$, $b_n = 0$ for $n \leq 0$, and $b_n$ is subharmonic w.r.t. $\pi$
Key lemma

Lemma
If $a_n, v_n \to 0$ and
1. $(v_n)_{n \geq 1}$ is harmonic w.r.t. policy $\pi$
2. $(a_n)_{n \geq 1}$ is subharmonic (superharmonic) w.r.t. $\pi$
3. $a_n = v_n$ for $n \leq 0$

Then $a_n \leq v_n$ ($a_n \geq v_n$) for all $n$.

Proof.

- Look at $b_n := a_n - v_n$
- $b_n \to 0$, $b_n = 0$ for $n \leq 0$, and $b_n$ is subharmonic w.r.t $\pi$
- If $b_n > 0$ for some $n$ then it achieves a maximum somewhere.
- The last maximum cannot be subharmonic.
Optimal strategies

Theorem

finite set of investments + irreducible investment A such that
\( \lambda_A^+ < \lambda_B^+ \) for all positive drift B’s \( \implies \) any optimal strategy has a
pure-A tail
Proof sketch

Idea: number of non-A’s in any optimal strategy is finite
Proof sketch

**Idea:** number of non-A’s in any optimal strategy is finite

$$\pi:$$ fixed strategy without a pure-A tail

$$\pi_A:$$ pure-A strategy with $$\lambda := \lambda_A$$

$$\nu_\pi:$$ insolvency probability using $$\pi$$

$$\nu_{\pi_A}:$$ insolvency probability using $$\pi_A$$
Proof sketch

Idea: number of non-A’s in any optimal strategy is finite

$\pi$: fixed strategy without a pure-A tail
$\pi_A$: pure-A strategy with $\lambda := \lambda_A$
$\nu^\pi$: insolvency probability using $\pi$
$\nu^{\pi_A}$: insolvency probability using $\pi_A$
$u^\pi$: harmonic w.r.t $\pi$, $u^\pi(w) \to 0$ as $w \to \infty$, $u^\pi(w) = \lambda^w$ for $w \leq 0$
$u^{\pi_A}$: $u^{\pi_A}(w) = \lambda^w$ for all $w$
Proof sketch

Idea: number of non-A’s in any optimal strategy is finite

π: fixed strategy without a pure-A tail
π_A: pure-A strategy with λ := λ_+
v_π: insolvency probability using π
v_π_A: insolvency probability using π_A
u_π: harmonic w.r.t π, u_π(w) → 0 as w → ∞, u_π(w) = λ^w for w ≤ 0
u_π_A: u_π_A(w) = λ^w for all w

\[ u_π_A \geq v_π_A \geq \lambda^l u_π_A \]
\[ u_π \geq v_π \geq \lambda^l u_π \]
**Proof sketch**

**Idea:** number of non-A’s in any optimal strategy is finite

\( \pi \): fixed strategy without a pure-A tail

\( \pi_A \): pure-A strategy with \( \lambda := \lambda^A_+ \)

\( v^\pi \): insolvency probability using \( \pi \)

\( v^{\pi_A} \): insolvency probability using \( \pi_A \)

\( u^\pi \): harmonic w.r.t \( \pi \), \( u^\pi(w) \to 0 \) as \( w \to \infty \), \( u^\pi(w) = \lambda^w \) for \( w \leq 0 \)

\( u^{\pi_A} \): \( u^{\pi_A}(w) = \lambda^w \) for all \( w \)

\[
\begin{align*}
    u^{\pi_A} & \geq v^{\pi_A} \geq \lambda^I u^{\pi_A} \\
    u^\pi & \geq v^\pi \geq \lambda^I u^\pi \\
    u^{\pi_A}(w) & < \lambda^I u^\pi(w) \implies \\
    v^{\pi_A}(w) & < v^\pi(w) \implies \pi \text{ cannot be optimal.}
\end{align*}
\]
Proof sketch

Starting with $u^{\pi_A} \equiv (\lambda^w)_w$ find a sequence that is subharmonic w.r.t $\pi$ (so that it is below $u^\pi$) yet satisfies the desired strict inequality.
Proof sketch

Starting with $u^{\pi^A} \equiv (\lambda^w)_w$ find a sequence that is subharmonic w.r.t $\pi$ (so that it is below $u^{\pi}$) yet satisfies the desired strict inequality.

▶ find a non-A investment
Proof sketch

Starting with $u^{π_A} ≡ (λ^w)_w$ find a sequence that is *subharmonic* w.r.t $π$ (so that it is below $u^π$) yet satisfies the desired strict inequality.

- find a non-A investment
- replace current value with its average; the new sequence is *strictly greater* there since $λ$ is the unique minimum $λ_+$
Proof sketch

Starting with $u^\pi_A \equiv (\lambda^w)_w$ find a sequence that is subharmonic w.r.t $\pi$ (so that it is below $u^\pi$) yet satisfies the desired strict inequality.

- find a non-A investment
- replace current value with its average; the new sequence is strictly greater there since $\lambda$ is the unique minimum $\lambda_+$
- replace “neighbors” by their (new) average: A being irreducible ensures that appropriate neighbors are increased
Proof sketch

Starting with $u^{\pi_A} \equiv (\lambda^w)_w$ find a sequence that is subharmonic w.r.t $\pi$ (so that it is below $u^{\pi}$) yet satisfies the desired strict inequality.

- find a non-A investment
- replace current value with its average; the new sequence is strictly greater there since $\lambda$ is the unique minimum $\lambda_+$
- replace “neighbors” by their (new) average: A being irreducible ensures that appropriate neighbors are increased
- repeat till the $(\frac{1}{\lambda})^l$ ratio is crossed (also gives a quantitative bound number of non-A’s in an optimal strategy)
Proof sketch

Starting with $u^{\pi_A} \equiv (\lambda^w)_w$ find a sequence that is subharmonic w.r.t $\pi$ (so that it is below $u^{\pi}$) yet satisfies the desired strict inequality.

- find a non-A investment
- replace current value with its average; the new sequence is strictly greater there since $\lambda$ is the unique minimum $\lambda_+$
- replace “neighbors” by their (new) average: A being irreducible ensures that appropriate neighbors are increased
- repeat till the $(\frac{1}{\lambda})^l$ ratio is crossed (also gives a quantitative bound number of non-A’s in an optimal strategy)
Proof sketch

Starting with $u^{\pi_A} \equiv (\lambda^w)_{w}$ find a sequence that is subharmonic w.r.t $\pi$ (so that it is below $u^{\pi}$) yet satisfies the desired strict inequality.

- find a non-A investment
- replace current value with its average; the new sequence is strictly greater there since $\lambda$ is the unique minimum $\lambda_+$
- replace “neighbors” by their (new) average: A being irreducible ensures that appropriate neighbors are increased
- repeat till the $(\frac{1}{\lambda})^l$ ratio is crossed (also gives a quantitative bound number of non-A’s in an optimal strategy)
Proof sketch

Starting with $u^{\pi_A} \equiv (\lambda^w)_w$ find a sequence that is *subharmonic* w.r.t $\pi$ (so that it is below $u^\pi$) yet satisfies the desired strict inequality.

- find a non-A investment
- replace current value with its average; the new sequence is *strictly greater* there since $\lambda$ is the unique minimum $\lambda_+$
- replace “neighbors” by their (new) average: A being irreducible ensures that appropriate neighbors are increased
- repeat till the $(\frac{1}{\lambda})^I$ ratio is crossed (also gives a quantitative bound number of non-A’s in an optimal strategy)
Proof sketch

Starting with \( u^{\pi_A} \equiv (\lambda^w)_w \) find a sequence that is \textit{subharmonic} w.r.t \( \pi \) (so that it is below \( u^\pi \)) yet satisfies the desired strict inequality.

- find a non-A investment
- replace current value with its average; the new sequence is \textit{strictly greater} there since \( \lambda \) is the unique minimum \( \lambda_+ \)
- replace “neighbors” by their (new) average: A being irreducible ensures that appropriate neighbors are increased
- repeat till the \((1/\lambda)^l\) ratio is crossed (also gives a quantitative bound number of non-A’s in an optimal strategy)
Non-pure tail

Theorem
\[ \lambda^A_+ = \lambda^B_+ < 1 \text{ and } A \text{ and } B \text{ have eigenvalues (in the unit disk) that are not shared by the other} \implies \text{any strategy with a pure-A (or pure-B) tail is not optimal} \]
Non-pure tail

**Theorem**

\[ \lambda^A_+ = \lambda^B_+ < 1 \] and A and B have eigenvalues (in the unit disk) that are not shared by the other \( \implies \) any strategy with a pure-A (or pure-B) tail is not optimal

**Main idea** \((l_A = l_B = 2)\).

- \( \lambda := \lambda^A_+ = \lambda^B_+ \) and the other eigenvalues \( \mu^A, \mu^B < 0 \)
- \( \pi \): strategy with pure-A tail
Non-pure tail

Theorem

\[ \lambda_+^A = \lambda_+^B < 1 \text{ and } A \text{ and } B \text{ have eigenvalues (in the unit disk) that are not shared by the other} \implies \text{any strategy with a pure-A (or pure-B) tail is not optimal} \]

Main idea \((l_A = l_B = 2)\).

- \(\lambda := \lambda_+^A = \lambda_+^B\) and the other eigenvalues \(\mu^A, \mu^B < 0\)
- \(\pi\): strategy with pure-A tail
- For some \(w > w_0\): 
  \[ v^\pi(w) = c_1 \lambda^{n-n_0} + c_2 (\mu^A)^{n-n_0} \]
- The first term is harmonic w.r.t. B too;
Non-pure tail

Theorem
\[ \lambda_+^A = \lambda_+^B < 1 \] and A and B have eigenvalues (in the unit disk) that are not shared by the other \( \implies \) any strategy with a pure-A (or pure-B) tail is not optimal

Main idea \((l_A = l_B = 2)\).

- \( \lambda := \lambda_+^A = \lambda_+^B \) and the other eigenvalues \( \mu^A, \mu^B < 0 \)
- \( \pi\): strategy with pure-A tail
- For some \( w > w_0 \): \( v^\pi(w) = c_1 \lambda^{n-n_0} + c_2(\mu^A)^{n-n_0} \)
- The first term is harmonic w.r.t B too;
- Two cases (both implying \( \pi \) not optimal)
  1. \( v^\pi \) is strictly superharmonic w.r.t. the strategy obtained by replacing \( \pi(w_0 + 3) \) by B, or
  2. \( v^\pi \) is strictly superharmonic w.r.t. the strategy obtained by replacing \( \pi(w_0 + 4) \) by B
Convergence rate for value iteration

\( \pi^* \): optimal policy
\( \nu^* \): optimal expected total value
\( \nu^n \): \( n \)th iterate of value iteration
\( u^n \): \( n \)th iterate of value iteration with action sets \( A_s = \{ \pi^*(s) \} \)
Convergence rate for value iteration

\( \pi^* \): optimal policy  
\( \nu^* \): optimal expected total value  
\( \nu^n \): \( n \)th iterate of value iteration  
\( u^n \): \( n \)th iterate of value iteration with action sets \( A_s = \{\pi^*(s)\} \)

\( u^n \rightarrow \nu^* \) and \( \nu^n \geq u^n \) (reminder: \( \nu \) is now the complement of the insolvency probability)
Convergence rate for value iteration

\( \pi^* \): optimal policy
\( \nu^* \): optimal expected total value
\( \nu^n \): \( n \)th iterate of value iteration
\( u^n \): \( n \)th iterate of value iteration with action sets \( A_s = \{ \pi^*(s) \} \)

\( u^n \to \nu^* \) and \( \nu^n \geq u^n \) (reminder: \( \nu \) is now the complement of the insolvency probability)

\[ u^n = Pu^{n-1} + c, \]

where, \ldots
The problem

Pure strategies

Pure strategies

Algorithms

Some details

\[ P = \begin{bmatrix}
p_{1,1} & \cdots & p_{1,M-1} & p_{1,M} & p_{1,M+1} & \cdots & p_{1,M+r-1} & p_{1,M+r} \\
\vdots & & \ddots & \vdots & & \vdots & \vdots & \vdots \\
p_{M,1} & \cdots & p_{M,M-1} & p_{M,M} & p_{M,M+1} & \cdots & p_{M,M+r-1} & p_{M,M+r} \\
\cdots & & \alpha_2 & \alpha_1 & 0 & 0 & \cdots & 0 \\
\cdots & & \alpha_2 & \alpha_1 & 0 & 0 & \cdots & 0 \\
\vdots & & \vdots & & \vdots & & \vdots & \vdots \\
\cdots & & 0 & 0 & \cdots & \alpha_1 & 0 & 0
\end{bmatrix} \]

bound on \( \rho(P) \)

\[ \sum_{j=1}^{M+r} p_{i,j} \leq 1; \sum_{j=1}^{M+r} \alpha_j < 1 \]

P need not be irreducible

Look at the Geršgorin region

if 1 is an eigenvalue then for some \( i \),

\[ \sum_{j=1}^{M+r} p_{i,j} = 1 \]

since this an optimal policy, there is positive probability of going from state \( i \) to the tail \( \{M+1, \ldots, M+r\} \) so that \( \sum_{j=1}^{M+r} \alpha_j = 1 \)
The problem

Pure strategies

Algorithms

Some details

\[ P = \begin{bmatrix}
  p_{1,1} & \cdots & p_{1,M-1} & p_{1,M} & p_{1,M+1} & \cdots & p_{1,M+r-1} & p_{1,M+r} \\
  \vdots & & & \vdots & & & \vdots & \\
  p_{M,1} & \cdots & p_{M,M-1} & p_{M,M} & p_{M,M+1} & \cdots & p_{M,M+r-1} & p_{M,M+r} \\
  \cdots & \alpha_2 & \alpha_1 & 0 & 0 & \cdots & 0 & 0 \\
  \cdots & \alpha_2 & \alpha_1 & 0 & \cdots & 0 & 0 & \\
  \vdots & & & \vdots & & & \\
  \cdots & 0 & 0 & \cdots & \alpha_1 & 0 & \\
\end{bmatrix} \]

- bound on \( \rho(P) \)

- \( \sum_{j=1}^{M+r} p_{i,j} \leq 1; \sum_{j=1}^{l} \alpha_j < 1; P \text{ need not be irreducible} \)
The problem

Pure strategies

Pure strategies

Algorithms

Some details

\[ P = \begin{bmatrix}
p_{1,1} & \cdots & p_{1,M-1} & p_{1,M} & p_{1,M+1} & \cdots & p_{1,M+r-1} & p_{1,M+r} \\
\vdots & & & & & & & \\
p_{M,1} & \cdots & p_{M,M-1} & p_{M,M} & p_{M,M+1} & \cdots & p_{M,M+r-1} & p_{M,M+r} \\
\cdots & \alpha_2 & \alpha_1 & 0 & 0 & \cdots & 0 & 0 \\
\cdots & \alpha_2 & \alpha_1 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & & & & & & & \\
\cdots & 0 & 0 & \cdots & \alpha_1 & 0
\end{bmatrix} \]

* bound on \( \rho(P) \)
* \( \sum_{j=1}^{M+r} p_{i,j} \leq 1; \sum_{j=1}^{l} \alpha_j < 1; P \text{ need not be irreducible} \)
* Look at the Geršgorin region
  * if 1 is an eigenvalue then for some \( i \), \( \sum_{j=1}^{M+r} p_{i,j} = 1 \)
The problem

Pure strategies

Pure strategies

Algorithms

Some details

\[ P = \begin{bmatrix}
  p_{1,1} & \cdots & p_{1,M-1} & p_{1,M} & p_{1,M+1} & \cdots & p_{1,M+r-1} & p_{1,M+r} \\
  \vdots & & \vdots & & \vdots & & \vdots & \\
  p_{M,1} & \cdots & p_{M,M-1} & p_{M,M} & p_{M,M+1} & \cdots & p_{M,M+r-1} & p_{M,M+r} \\
  \cdots & \alpha_2 & \alpha_1 & 0 & 0 & \cdots & 0 & 0 \\
  \cdots & \alpha_2 & \alpha_1 & 0 & 0 & \cdots & 0 & 0 \\
  \vdots & & \vdots & & \vdots & & \vdots & \\
  \cdots & 0 & 0 & \cdots & \alpha_1 & 0 & \\
\end{bmatrix} \]

- bound on \( \rho(P) \)
- \( \sum_{j=1}^{M+r} p_{i,j} \leq 1; \sum_{j=1}^{l} \alpha_j < 1; \) \( P \) need not be irreducible
- Look at the Geršgorin region
  - if 1 is an eigenvalue then for some \( i, \sum_{j=1}^{M+r} p_{i,j} = 1 \)
  - since this an optimal policy, there is positive probability of going from state \( i \) to the tail \( \{M+1, \ldots, M+r\} \) so that \( \sum_{j=1}^{l} \alpha_j = 1 \)
Bound on spectral radius for positive drifts

\( D \): diagonal matrix with entries \((d_1, \ldots, d_{M+r}) > 0\)

Then \( \rho(P) = \rho(D^{-1}PD) \leq \|D^{-1}PD\|_\infty \)
Bound on spectral radius for positive drifts

\( D \): diagonal matrix with entries \((d_1, \ldots, d_{M+r}) > 0\)

Then \( \rho(P) = \rho(D^{-1}PD) \leq \|D^{-1}PD\|_\infty \)

Choose \( d_i = \mu^i \). The \( i \)th row sum (corresponding to investment B) of \( D^{-1}PD \) is then (for \( 1 \leq i \leq M \)),

\[
\sum_{j=1}^{M+r} p_{i,j} \frac{d_j}{d_i} = \sum_{j=1}^{M+r} p_{i,j} \mu^{j-i} \leq 1 - \frac{q^B(\mu)}{\mu^B} := \delta_B
\]

If \( \lambda_+^B < \mu < 1 \) then \( \delta_B < 1 \).
Bound on spectral radius for positive drifts

$D$: diagonal matrix with entries $(d_1, \ldots, d_{M+r}) > 0$

Then $\rho(P) = \rho(D^{-1}PD) \leq \|D^{-1}PD\|_{\infty}$

Choose $d_i = \mu_i$. The $i$th row sum (corresponding to investment $B$) of $D^{-1}PD$ is then (for $1 \leq i \leq M$),

$$\sum_{j=1}^{M+r} p_{i,j} \frac{d_j}{d_i} = \sum_{j=1}^{M+r} p_{i,j} \mu^j_i \leq 1 - \frac{q^B(\mu)}{\mu_B} := \delta_B$$

If $\lambda^B_+ < \mu < 1$ then $\delta_B < 1$.

- similar argument for $M + 1 \leq i \leq M + r$
- Choose $\mu$ to be larger than all the $\lambda_+$'s
- $\|D^{-1}PD\|_{\infty}$ is the maximum of all the $\delta$'s